Gait recognition: Monocular, RGB-D, Appearance and Model based methods

University of Patras Computer Vision Group,
Department of Physics,
Electronics laboratory

Dimitris Kastaniotis, PhD Candidate
Contents

• About Gait
  – Gait in Cognitive Neuroscience
  – Gait attributes

• Data describing gait (RGB, Depth, Motion Capture Data) - Features and feature representation of gait

• Appearance and Pose based methods for Gait recognition
  – Monocular (RGB)
  – Monocular Depth

• Depth Imaging and Pose estimation
  – Depth Imaging in Gait Analysis
  – Kinect Pose estimation algorithm

• Applications
Introduction

• Gait is a human motion
• Encompasses meaningful Biological information
• Gait is unique\(^1\)
• Works at a distance
• Unobtrusive
• Confidential
• Hard to imitate

Introduction

• How people understand action-movement
• Johansson et. al. 1971 proposed Point Light Walkers

[M.A. Giese & Tomaso Poggio, Nature Reviews Neuroscience 4, 179-192 (March 2003)]
Introduction

• How people understand action-movement
• Johansson et. al. 1971 proposed Point Light Walkers (PLW)
  • Biological motion cues
  • People can recognize gender in PLW with 65–75% correct classification rate

Source: http://www.youtube.com/watch?v=r0kLC-priID
Introduction

• **How does this work?**

• **Motion pathway.** The motion pathway recognizes, biological movements by analyzing optic-flow patterns.

• **Form pathway.** The form pathway analyses biological, movements by recognizing sequences of ‘snapshots’ of body shapes.
Introduction

- IT- inferotemporal cortex;
- KO- kinetic occipital cortex
- OF- optic flow
- RF- receptive field
- STS- superior temporal sulcus
- V1- primary visual cortex
- V2- Visual area V2, called prestriate cortex

[M.A. Giese & Tomaso Poggio, 2003]
Summary of Gait “Attributes”

• Given a gait sequence a person is able to:
  – Recognize a familiar person
  – Recognize the gender
  – Estimate the mental state of a person
  – Estimate the weight
  – Provide medical assessment
  – Diagnosis and/or treatment of gait-related disorders

Elicit information from a given gait sequence in order to perform different tasks

[Jinfu Shi et.al. Cognition 2010]
A System for Gait Recognition

• Human Visual system efficiently processes visual information

• In order for a system to achieve the same or better performance:
  – What kind of data should make use of?
  – What are the most appropriate features describing those data?
  – How many frames required? ¹
  – How recognition is performed?

¹Konrad Schindler, Luc Van Gool, “Action Snippets: How many frames does human action recognition require?”
A “Complete” System for Gait Recognition

• Capture data
• Extract features
• Train a classifier to **learn** how to perform a **TASK**

... then...

• Given a “**test**” sample, an appropriate system should be able to **classify** it into the correct class
Gait Data Analysis

- Different nature of source data
- Different feature representation
- Different task {gender, identification, age, mental state, medical diagnosis, etc.}
Gait Recognition Pipeline

• **Acquire** some data

• **Preprocess** and Represent Data conveniently

• Choose a **Feature Representation** of data

• **Classify**
  – K-nn, SVM, Neural Networks, Sparse Based Classification
Data Acquisition, Feature Extraction and Classification

– Monocular
– RGB-D Sensors

Representation
Once data are captured, there are several ways to represent them:

**Mode Free**

**Model Based**
– Top Down (compare a 3D model with current image)
– Bottom Up (assembling detected parts)

Features

**Static** \{*stride, stance, cadence, height, etc*\}

**Dynamic** \{*thighs joint angles, etc*\}

**Fusion of Static and Dynamic**

Classification
SVM, Neural Networks, KNN, SRC

[R. Poppe, 2010]
Overview

Monocular
- Appearance Based
  - Silhouette
  - Contour
- Model Based
  - Pose estimation

RGB-D
- Appearance
- Depth Silhouette
- Model Based
- Skeleton
Overview

Monocular

- Appearance Based
  - Silhouette
  - Contour

- Model Based
  - Pose estimation

RGB-D

- Appearance
- Depth Silhouette
- Model Based
- Skeleton
Gait Based Recognition: Appearance Vs Model Based

Model Based
- High Level Processing
- Use of static and dynamic body parameters which are general scale and view invariant
- high accuracy but **requires** high quality images

Model Based also called Pose based

Robust to interclass variances

Model Free
- High Level Processing
- Very low resolutions
- Low Computational Costs
- Low Time Costs

• **Non Robust to interclass variances**
Roadmap

Monocular

- Appearance Based
  - Silhouette
  - Contour
- Model Based
  - Pose estimation
Roadmap

Monocular

Appearance Based
- Silhouette
- Contour

Model Based
- Pose estimation
Gait Based Recognition
Model Free Approaches

• Gait Energy Image [Ju Han et. al. 2006, PAMI]

  Averages information
  - Time information is lost
  - Sensitive to carrying objects
  - Scale and View dependent

  [Ju Han et. al. 2006, PAMI]

\[ G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B_t(x, y), \]

Also Motion History images have been applied in Gait analysis...
Gait Based Recognition: Monocular Appearance Based

- Gait Histogram Energy Images GHEI \(^1\)
- Builds on the Gait Energy Image + HOG \(^2\)
- Doesn’t average silhouettes.
- Uses orientation histograms

GEHI is obtained by averaging the gradient histogram Representations over a full gait cycle

Variants:
- s- GEHI: Coarse Foreground Segmentation
- a- GEHI:k

PCA+LDA, classification by Nearest Neighbor

\(^1\)Hofmann et. al. 2012, \(^2\)[Dalal & Trigs, 2005]
Roadmap

Monocular

- Appearance Based
  - Silhouette
  - Contour
- Model Based
  - Pose estimation
Roadmap

Monocular

Appearance Based

Silhouette

Contour

Model Based

Pose estimation
Gait Based Recognition: Monocular Appearance Based I

Chrono- Gait Image

As we saw before: “Gait energy images discard temporal information”

- Essentially is a “template” for encoding temporal information in a single color image

Steps:
- Contour extraction
- Color coding temporal information
- Applies PCA+ LDA
- Matching performed using the nearest neighbor distance w.r.t. some templates

GCI Template: \( CG(x, y) = \frac{1}{p} \sum_{i=1}^{p} PG_i(x, y) \), where \( p \) is period

[Cheng et.al. ECCV 2010] Cheng et.al. PAMI 2012]
Roadmap

Monocular

- Appearance Based
  - Silhouette
  - Contour
- Model Based
  - Pose estimation
Roadmap

Monocular

- Appearance Based
  - Silhouette
  - Contour

- Model Based
  - Pose estimation
Gait Based Recognition: Monocular Model Based

Using a 2D stick figure Representation

1. detection and extraction of the moving human body and its contour from image sequences
2. extraction of human gait signature by the joint angles and body points
3. motion analysis and feature extraction for classifying gender in the gait patterns.

Gender Classification in Human Gait Using Support Vector Machine, Jang-Hee Yoo, Doosung Hwang, and Mark S. Nixon
Gait Based Recognition: Monocular Model Based I

14 rigid parts, connected with Joints

- Used static and dynamic features

Bobick and Johnson “static body and stride”

Tanawongsuwan & Bobick “joint trajectories ~ MOCAP”

[Wang et al, 2004], [Bobick & Johnson], [Tanawongsuwan & Bobick]
Roadmap

RGB-D

- Appearance
- Depth Silhouette

- Model Based
- Skeleton
About Depth Imaging

• Reached a consumer price point

• Every pixel represent the calibrated distance in meters from sensor (no color or texture)

• (MS Kinect Sensor) Works with structured infra-red at 30 fps
Let's Make with a Quiz.....
QUIZ!

- You will watch two gait sequences.
  - Each person gait sequence is captured with depth video

- Video 1-TASK- A
  - Do you recognize the person walking?

- Video 2-TASK- B
  - Can you recognize its gender?
QUIZ!

• a. Guess Who
QUIZ -2

b. Can you recognize the gender of the person?
About Depth Imaging

• Applications in:
  – Gesture recognition
  – Rehabilitation (stroke patients)
  – Graphics Animation
  – Surveillance
  – Statistics from extracted attributes
  – Object Detection
RGB-D Appearance based methods

• 2.5D Gait Biometrics using the Depth Gradient Histogram Energy Image, BTAS 2012
  – Used Depth information to extract silhouette
  – Adaptes GEHI to "2.5D"

![Average gradient histograms over a full gait cycle](image)

$$H(i, j, f) = \frac{1}{T} \sum_{t=1}^{T} h_t(i, j, f)$$

$$L_i = \arg \min_c D_i(c)$$
RGB-D Appearance based methods

- Depth Information in Human Gait Analysis: An Experimental Study on Gender Recognition

Euclidean distances between pairs of randomly selected points on the border of the 2D silhouette or 3D surface

Classification using a Kernel SVM.

Available at: http://www.cvc.uab.es/DGaitDB/

[Ricard Borras et. a.]
RGB-D Appearance based methods

- Gait Energy Volumes and Frontal Gait Recognition using Depth Images
  - **Adapts** GEI to 3D (depth)

  Collected a database of 15 subjects walking towards the camera
  Two different speeds, ‘normal’ and ‘fast’.
  Microsoft Kinect

  Sivapalan, Sabesan et. al. 2010

\[
\text{GEV}(k) = \frac{1}{n} \sum_{t=1}^{n} V(t),
\]
RGB-D Appearance based methods

• The above methods are explicit extensions of monocular proposed ones.

• Are not appropriate in modeling kinematic dynamics
Roadmap

RGB-D

- Appearance
- Depth Silhouette

- Model Based
- Skeleton
Roadmap

RGB-D

- Appearance
- Depth Silhouette

- Model Based
- Skeleton
A question ... 

Does Human Action Recognition Benefit from Pose Estimation?

• Yes indeed…

Figure 5: Confusion matrices for appearance-based, pose-based and combined features for action recognition, with mean classification rates of 0.698, 0.815 and 0.801 respectively. While “release grasp” is the most confused action in all cases, classes like “take object” or “idle/carry (still)” significantly improve with the introduction of pose-based features.
OTHER pose estimation methods

• In order to detect and identify body parts in depth images at video frame rates.

They proposed interest points termed: AGEX, Accumulative Geodesic Extrema + Local Descriptor Patches

Requires computation of geodesic extrema on mesh

OTHER pose estimation methods

- Estimating Human 3D Pose from Time-of-Flight Images Based on Geodesic Distances and Optical Flow

Anatomical landmarks remain in constant geodesic distance from Center of mass

Loren Arthur Schwarz et al. 2011
Kinect Pose estimation Algorithm

• Requirements:
  – Skeleton Tracking algorithm
  – No pose initialization
  – Frame by Frame (no temporal information)
  – Recovering from failures
  – Real time (decision trees- parallelization)
  – Easy to compute features
  – Pixel level Classification
  – Robust to occlusions

[Shotton et. al. 2011, CVPR ,Shotton et. al. 2011, CVPR]
Kinect Pose estimation Algorithm

• Pose Estimation from Depth Images
  – No need for pose initialization
  – Very Fast on execution time
  – Parallelizable (on CPU/ GPU)

• Depth Images have several unique properties
• No lighting variances
• No color
• No texture

Kinect team took the advantage on them
Important Concept!

• **Classification** into body parts or **regression of offset vectors**?

• Imagine you have given a depth image.

• Someone asks you to fit a “skeleton” model on it using only depth data.

[Shotton et. al. 2011, CVPR, Shotton et. al. 2011, CVPR]
Important Concept!

• Classification in Body Parts or Regression?

[Shotton et. al. 2011, CVPR, Shotton et. al. 2011, CVPR]
Kinect Pose estimation Algorithms

• What kind of data BPC and OJR use?
  – Both Algorithms use only Depth (try it at home..)

• Feature Extraction
  – Both algorithms use the same features

BUT DIFFERENT LABELS

• Classification- Based on Ensemble of Randomized Decision Trees
  – BPC: categorical
  – OJR: continuous

[Shotton et. al. 2011, CVPR ,Shotton et. al. 2011, CVPR]
Feature Extraction

Each feature performs two offset ‘depth probes’

Feature parameters:
\[ \phi = (u, v) \] are 2D pixel offsets

\[ 1/d(x) \] is the distance of pixel “x” from sensor.

Depth normalization

\[ u, v \] are Bounded in some range

Offset vector \( v \) is zero with probability 0.5

A Generic Model for Kinect Algorithms

Step 1: Given a Depth image

Step 2: Features are extracted

Step 3: Classifier (Randomized Decision Trees) are trained
Kinect Pose estimation Algorithms

• Pose Estimation from Depth Images Using Kinect Core Algorithm
• Two basic Algorithms Proposed

• a. **Body Part Classification** - BPC
  – Every pixel is classified into one of “N” body parts
  – Predicts points on surface
  – Categorical Classification \{left hand, right hand, head, etc…\}

• b. **Offset Joint Regression** - OJC
  – Every pixel votes directly for the position of the different body joints
  – Directly predicts joints
  – Joints for invisible parts are successfully estimated

[Shotton et. al. 2011, CVPR ,Shotton et. al. 2011, CVPR]
Kinect Pose estimation Algorithms

Body Part Classification - BPC

– Classify every pixel in category {right hand, head, etc}

Mean Shift for Mode Detection

\[ f_\theta(I, x) = d_I \left( x + \frac{\mathbf{u}}{d_I(x)} \right) - d_I \left( x + \frac{\mathbf{v}}{d_I(x)} \right) \]

Offset Joint Regression- OJR

- Every pixel “votes” for the position of the joint
  - 3D offset votes
  - Defines a world space density
  - Using mean shift to find the final joint proposals

- Even occluded joints can be estimated

- Reduced error due to elimination of the intermediate step of clustering presented in BPC

Kinect Pose estimation Algorithm

• Randomized decision trees
  – **Multiclass** classification
  – Randomization in features or splitting function ensures good generalization
  – Very slow training
  – Very fast execution time (Parallelizable)

\[ P(c|I, x) = \frac{1}{T} \sum_{t=1}^{T} P_t(c|I, x). \]
Kinect Pose estimation Algorithm

- Randomized decision trees

![Graph showing average per-class accuracy vs number of trees]

- [J. Shotton, et. al. 2011]
- [J. Taylor, J. Shotton, CVPR 2012]
Kinect Pose estimation Algorithm

• Randomized decision trees *in categorization and segmentation*

Matthew Johnson and Jamie Shotton, *Semantic Texton Forests*, in Computer Vision: Recognition, Registration and Reconstructions, Springer Verlag, March 2010


6 channels: CT1 post gadolinium (T1-gad), T1, T2 turbo spin echo (T2-tse), and FLAIR, and 2 channels from diffusion tensor imaging (DTI-p and DTI-q).
Kinect Pose estimation Algorithm

- Kinect Algorithm Evolution with time
- Classify pixels into body parts
- Pixels vote for the position of joints
- Incorporate joint dependencies
- Vitruvian Manifold Model
  - Each pixel gives an estimate of the correspondence to an articulated mesh model

Kinect Gait Applications – I

• “Gait Recognition with Kinect”
• Features:
  • Height,
  • Length of legs,
  • Torso
  • Lower legs
  • both thighs
  • both upper arms
  • both forearms
  • Step length
  • Speed

[J. Preis et.al.]
Unsupervised Clustering of People from ‘Skeleton’ Data

- Used only lower limbs joint angles
- Mean, Standard Deviation and Maximum Values of several angles
- 18 total Features

K-mean clustering

Table 1: Clustering confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Label A</th>
<th>Label B</th>
<th>Label C</th>
<th>Label D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>35%</td>
<td>46%</td>
<td>8%</td>
<td>12%</td>
</tr>
<tr>
<td>Person 2</td>
<td>0%</td>
<td>74%</td>
<td>16%</td>
<td>11%</td>
</tr>
<tr>
<td>Person 3</td>
<td>21%</td>
<td>21%</td>
<td>29%</td>
<td>29%</td>
</tr>
<tr>
<td>Person 4</td>
<td>17%</td>
<td>17%</td>
<td>33%</td>
<td>33%</td>
</tr>
</tbody>
</table>

[Ball et.al.]
Gait based identification and gender recognition

Live Demo Application!
Gender Recognition using Kinect

• First Results

• 7 subjects

• Pairwise Distance Matrix

• Classic MDS
Other Applications: Home Monitoring

Passive In-Home Measurement of Stride-to-Stride Gait Variability Comparing Vision and Kinect Sensing

- 25-35% of people 65 years or older fall each year
- Additionally, research has identified specific measures of gait which may be predictive of future falls in older adults
- Analyzing gait variability (stride to stride may infer meaningful information)

[Erik E. Stone, & Marjorie Skubic, 2011]
A final Quiz...

You will see two gait sequences of the same Person in two different “states”.

What do you think really changed?
Find the differences...
References

Gait analysis

• Xuelong Li, Stephen J. Maybank, Shuicheng Yan, Dacheng Tao, and Dong Xu, Gait Components and Their Application to Gender Recognition, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 38, NO. 2, MARCH 2008
• Kaiqi Huang, Kui Jia, and Xinyu Wu, A Study on Gait-Based Gender Classification, Shiqi Yu, Tieniu Tan, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 8, AUGUST 2009 1905
• Jiwen Lu, and Yap-Peng Tan, Gait-Based Human Age Estimation, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 4, DECEMBER 2010
References

Kinect Pose Estimation

• Jamie Shotton, Ross Girshick, Andrew Fitzgibbon, Toby Sharp, Mat Cook, Mark Finocchio, Richard Moore, Pushmeet Kohli, Antonio Criminisi, Alex Kipman, and Andrew Blake, Efficient Human Pose Estimation from Single Depth Images, in Trans. PAMI, IEEE, 2012

• Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake, Real-Time Human Pose Recognition in Parts from a Single Depth Image, in CVPR, IEEE, June 2011

• Min Sun, Pushmeet Kohli, and Jamie Shotton, Conditional Regression Forests for Human Pose Estimation, in Proc. CVPR, IEEE, 2012

• Jonathan Taylor, Jamie Shotton, Toby Sharp, and Andrew Fitzgibbon, The Vitruvian Manifold: Inferring Dense Correspondences for One-Shot Human Pose Estimation, in Proc. CVPR, IEEE, June 2012
Cognitive Neuroscience and Biological Motion


• Jinfu Shi, Xuchu Weng, Sheng He, Yi Jiang, Biological motion cues trigger reflexive attentional orienting, Cognition 117 (2010) 348–354

• Joachim Lange and Markus Lappe, A Model of Biological Motion Perception from Configural Form Cues, The Journal of Neuroscience, March 15, 2006 • 26(11):2894 –2906
References

Surveys


“Thank You”

Questions

dkastaniotis@upatras.gr
University of Patras Computer Vision Group - UPCV
http://www.upcv.upatras.gr/
Homepage
http://www.upcv.upatras.gr/personal/kastaniotis